

Automatic Cell Counting using Active Deep Learning and Unbiased Stereology

Saeed S. Alahmari¹, Dmitry Goldgof¹, Lawrence O. Hall¹, and Peter R. Mouton^{1,2}

¹Department of Computer Science and Engineering, University of South Florida, Tampa, FL, USA

²SRC Bioscience, Tampa, FL USA

{saeed3,goldgof,lohall}@mail.usf.edu, and {peter}@disector.com

Abstract—Training supervised algorithms such as deep learning requires a large labeled datasets; however, labeled data are not enough for training deep learning, and labeling process is tedious, time-consuming, and requires expert knowledge. In this paper, we proposed an active deep learning approach for cell segmentation and counting based on unbiased stereology. This method allow obtaining labeled data with minimal human intervention by using snapshot ensemble to get a confidence score for each mask in an unlabeled pool, where the user verifies only the images of higher confidence. The proposed method showed an error rate of less than 1% in the unbiased stereology cell count of stained sections with about 25% reduction in a human-in-the-loop verification time cost compared to the previously proposed method (iterative deep learning).

I. INTRODUCTION

Understanding the behavior and the presence of certain diseases such as Alzheimer and cancer requires quantification of the total number of cells during the diagnostic process and treatment process. An approach of such quantification is called unbiased stereology, which is a set of theoretical and practical methods for making accurate counts of stained cells by carefully avoiding all known sources of methodological bias [1][2]. Examples of common stereology parameters include counts of total cell number and cell density; region and mean cell volumes; surface area and surface density; and total length and length density [3][1]. However, current computer-assisted stereology systems available to bioscientists and medical scientists are based on a technology developed more than two decades ago. Though based on theoretically unbiased principles, this approach is prone to data errors and low reproducibility due to user subjectivity, variable expertise, and fatigue. The adaptive Segmentation algorithm was proposed in [2] to automate cell count in stained sections. Moreover, a deep learning approach was proposed in [4] to automate cell segmentation and counting based on unbiased stereology.

The Adaptive Segmentation Algorithm (ASA) [2] makes stereology counts of total numbers of brain cells (NeuN immunostained neurons) by automatic segmentation and cell counting on Extended Depth of Field (EDF) images [5][6]. However, ASA requires manual adjustment of several parameters (i.e., minimum cell size, cell maximum size, and Gaussian Mixture Model (GMM) threshold) to achieve a good result. In Section IV, we present ASA details.

Deep neural networks have generated considerable interest

in the medical imaging field because they have shown performance advantages over conventional engineered image analysis algorithms. Although the idea of neural networks has been around for a long time, the recent deep neural networks revolution is partly due to the development of the convolutional neural network (CNN), optimization algorithms [7] [8] [9] [10], and powerful, efficient computation resources. Deep learning refers to learning methods that often start from raw data get to a more abstract level [11]. Convolutional Neural Networks have shown significant success in challenging tasks in image classification and recognition [12] [13]. In this paper, we use a CNN based architecture for medical image segmentation known as Unet [14]. This architecture is a simple, fast, and end-to-end fully convolutional network that contains contraction and expansion paths to capture context and learn precise localization.

Supervised learning algorithms such as deep learning requires an extensive labeled data to learn from; however, labeled data is inadequate for many deep learning applications such as medical images analysis, because data labeling is time-consuming, expensive, and labor intensive [15]. Active learning is an artificial intelligence technique that solves the labeling bottleneck by querying the most uncertain instances of unlabeled data to be labeled by a user and added to the training instances [15]. This technique has been used in many applications such as image retrieval [16], support vector machine based text classification [17], gene expression classification [18], and interactive image segmentation [19][20][21]. However, labeling medical image for segmentation (i.e., creating a pixel-level label) is hard, time-consuming, and requires medical expert knowledge. Therefore, in this paper, we followed a slightly different approach when querying a pool of unlabeled data, so the confidence is calculated using snapshot ensemble of deep learning models, and the user verifies the most certain masks. Verification was done by human-in-the-loop where just accepting or rejecting the of masks are performed, and no manual labeling is involved as explained in Section VI.

Snapshot ensemble approach was proposed in [22], which allow obtaining multiple deep learning models saved during a single deep neural network training, and eliminates any extra-training cost of multiple neural networks individually. Snapshot ensemble takes advantage of training convergence of multiple local minima, and it uses a cyclic learning rate

scheduling technique to make performance variation in the saved snapshot models.

In this paper, we propose a method based on deep learning with snapshot ensemble to enable time reduction of the human-in-the-loop verification. This method uses an existing unsupervised algorithm (ASA) for initial data labeling of NeuN stained images to quantify the number of cells in an ROI. This approach uses a state-of-art deep learning architecture in which user verified ASA results of EDF images are used to train a convolution neural network (CNN) model to segment and make automatic neuron counts on test images. Meanwhile, a set of deep learning predicted masks of high confidence scores are verified by a human-in-the-loop and fed back to the train set. The main innovation is: i) reduction of human-in-the-loop verification time compared to [23] by using verifying only images of high confidence calculated using snapshot ensemble, ii) improving deep learning stereology cell counting by adding correctly labeled images (EDF images and their corresponding masks) to the training set for the next iteration.

II. UNBIASED STEREOLOGY

Unbiased stereology is the state-of-the-art for biological objects quantification in tissue sections [24]. An essential component of this approach is unbiased sampling (i.e., systematic-random) that avoids all sources of biased assumption such as shape, size, and orientation [24] [3]. Unbiased stereology uses a virtual disector box to quantify the number of cells in a region-of-interest (ROI). Counted cells are based on their location within an ROI and disector box. For instance, cells touching the disector inclusion-line (i.e., disector upper and right line) or inside the disector box are counted. However, cells that touch the exclusion line (i.e., disector lower and left line) are not counted. An example of the disector box counting procedure is shown in Fig. 1a, where the green line represents inclusion line, and the red line represents the exclusion line. Counted cells are marked manually with the blue marks.

III. DATA SET

The data set used in this work was sampled from the neocortex brain region of Tg4510 mice. As described by Mouton et al. in [2], animals and the process used in this study were approved by the University of South Florida (USF) Institutional Animal Care and Use Committee which follows NIH guidelines. The data set includes both genetically modified mice and control mice. Mice neurons change while expressing mutant tau. These neuron changes include neuron degeneration and neuroglia cells activation [2][25][26]. Mice samples were stained with NeuN single staining from which counting was performed manually using an optical fractionator [27]. Disector stacks were captured and saved using the Stereologer system [2]. Table I shows the number of sections from which multiple stacks were obtained and converted into EDF images. The total number of EDF images we have is 966 with their corresponding ASA masks.

TABLE I
DATASETS MOUSE ID, NUMBER OF SECTIONS PER MOUSE AND TOTAL
NUMBER OF STACKS PER MOUSE

Mouse ID	Number of sections	Number of stacks
02	8	113
03	6	121
14	8	90
17	7	91
29	8	135
21	7	102
24	8	103
67	8	104
09	6	107

IV. ADAPTIVE SEGMENTATION ALGORITHM

As shown in [2], the adaptive segmentation algorithm (ASA) consists of multiple steps optimized to segment cells at high magnification (63 to 100x oil immersion) microscopy. The ASA includes a Gaussian Mixture (GMM), morphological operations, Voronoi diagrams, and watershed segmentation. It starts with EDF images to segment NeuN stained cells within a region of interest (ROI) using GMM; where GMM uses pixel intensity for the Expectation Maximization algorithm (EM) to estimate its components followed by thresholding and morphological operations to get separate cells. A processed EDF image using opening then closing by reconstruction is used in the watershed foreground and background markers extraction. These foreground and background markers are used in applying the watershed segmentation followed by segmentation approximation using Voronoi diagrams. ASA uses a smoothing process to enhance cell boundaries using a Savitzky-Golay filter [28]. The reason to use ASA is that manual annotation does not provide mask information, but instead, it provides a mark of what cell is being counted based on the unbiased stereology approach. An example of manual annotation is shown in Fig. 1a.

V. ITERATIVE DEEP LEARNING

Iterative deep learning algorithm was proposed in [23], to enhance deep learning models segmentation performance by increasing the training data examples in an iterative approach. This method uses a pool of unlabeled data called active set, where a deep learning model is trained on a training set and then used to predict (i.e., generate masks) on the active set of EDF images. After that, human-in-the-loop verifies all predicted masks generated for all available EDF images in the active set (i.e., verifying the agreement of a mask and the corresponding manual annotation). If a mask is accepted, then the mask and EDF is augmented and added to the training set; otherwise, it is rejected. This process was performed for 5 iterations, and the five resulting models from all iterations were used to predict (i.e., generate masks) on a test set image, followed by post-processing and counting. Although this process is effective in increasing the number of training images by utilizing previously trained model to generate labels (i.e., masks) for a pool of unlabeled set (i.e., active set), the time consumed by the human in the

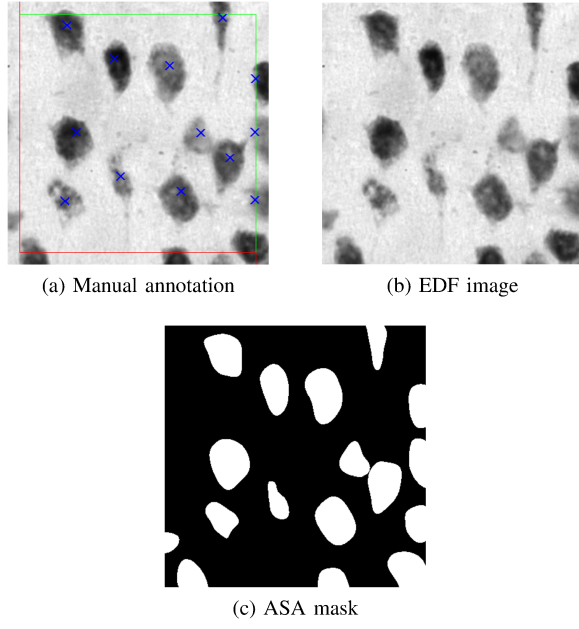


Fig. 1. An example from our data set, where a) is the manual annotation (counted neurons have green dots), b) is the EDF image, and c) is the ASA mask for the EDF image shown in (b)

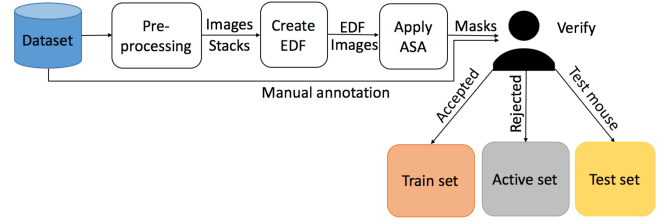
verification step is substantial and requires huge human effort especially for large unlabeled sets, and dense cells images. Therefore, utilizing the information from predicting masks using an ensemble approach to derive a confidence score that can be used to identify the most confidence predicted masks using the neural network would help the user to get verification of images done quicker.

VI. METHOD

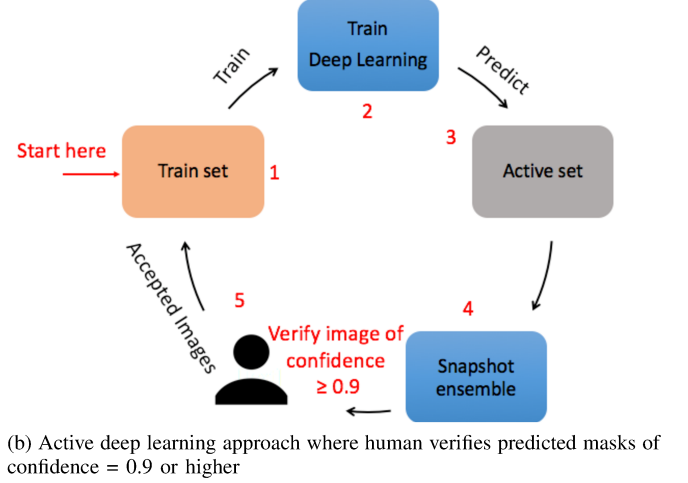
The main drawback of iterative deep learning [23] is that the human-in-the-loop needs to verify all deep learning models' predicted masks on the active set (i.e., ASA based rejected EDF images), which is time-consuming especially for a large set of unlabeled images. Moreover, the decision on how good is a predicted mask is a hard problem and requires a human to verify the goodness of a predicted mask. However, the human ability to verify a large number of unlabeled sets is limited. Therefore, to reduce the verification effort, it is crucial to utilize information from deep learning predicted masks which can alleviate the burden of verifying every single image/mask. In this paper, we propose an active deep learning approach to reduce verification effort by reducing the amount of time taken by a human-in-the-loop to verify the predicted masks. This approach is based on snapshot ensembles to measure the confidence of every predicted mask. Then human get only the images of confidence score ≥ 0.9 . Active deep learning is in three steps as follows: i) data preparation, ii) training active deep learning, and iii) testing process and post-processing

A. Data Preparation

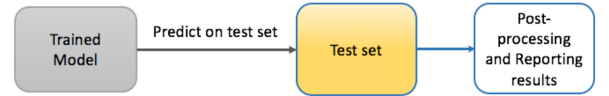
Data set of 966 stacks of microscopy images (as described in Section III) were pre-processed to convert each image in a



(a) Initial masks created using ASA followed by human verification



(b) Active deep learning approach where human verifies predicted masks of confidence = 0.9 or higher



(c) Testing and post-processing step

Fig. 2. Proposed method in two steps: a) pre-processing stacks, creating EDF images, and applying ASA, then human verification, and then b) active deep learning process using accepted ASA masks/images for training, and ASA masks/images as an active set. Human verification (i.e., accept or reject) on subset of predicted mask based on confidence. Test set is a separate mouse (mouse id 17)

stack from color to gray, then we applied the EDF algorithm to create a single synthetic image of each stack where all cells are in-focus as shown in Fig. 1b. After that, we applied the ASA algorithm to segment cells on EDF images (ASA is described in Section IV. ASA mask is shown in Fig. 1c. Then a human verification process is applied, where a human verifies the agreement of an ASA mask and the corresponding manual annotation as shown in Fig. 2a. If an ASA mask matches the manual annotation (i.e., every counted cell in manual annotation has the corresponding binary mask in the ASA mask), then it is accepted and moved to train set. If ASA mask does not match manual annotation (i.e., at least one marked cell in the manual annotation image does not have the corresponding mask), then it is rejected, and the EDF image is moved to the Active set. A separate mouse was chosen randomly as a test set for which ground truth was corrected manual.

B. Training Active Deep Learning

Our Active deep learning approach uses the initial training set generated during data preparation, where the labels of EDF images (i.e., segmentation masks) are based on ASA verification process as shown in Fig. 2a. The Active set has EDF images from which ASA masks were rejected during the ASA verification process in the data preparation process. It is important to note that the Active set has no associated labels (masks), and thus, a deep learning model in each iteration is used to generate masks (i.e., predict on the Active set). The process of active deep learning is as follows: 1) train deep learning model, 2) predict on active set, 3) compute confidence, 4) human verification.

1) *Train deep learning model:* We trained deep learning model using an off-shelf deep neural network called Unet [14] for 100 epochs using Keras and Tensorflow deep learning frameworks [29][30]. The Adam optimizer was used where the learning rate was set to $1e^{-4}$, while exponential decay rates for the moment estimates hyper-parameters β_1 (first moment) and β_2 (second moment) were set to 0.9 and 0.999 respectively [31]. During training, we used a snapshot model saving approach, where saving snapshots model every specific number of epochs is done, to account for variability of knowledge learned during training. In our experiment, snapshots models were saved every 5 epochs starting from epoch 10 to epoch 100. Total number of snapshots models are 19 models $\{M_1, M_2, \dots, M_{19}\}$.

2) *Predict on active set and apply snapshot ensemble:* After completing training of a deep learning model and saving the snapshots models, we used each snapshot model $\{M_1, M_2, \dots, M_{19}\}$ to predict on the active set Z , the results of prediction on an EDF image Z_{EDF} is a probability map Z_{prob} , where each pixel $p \in (0, 1)$. We thresholded the probabilities map Z_{prob} at 0.5, such that each pixel $(p > 0.5) = 1$ (i.e., foreground), otherwise $p = 0$ (i.e., background). Let the thresholded probability map be Z_{thrsh} . After thresholding, the results of using snapshot model M_i where $i \in \{1, 2, \dots, 19\}$ on the active set is averaged $\frac{1}{19} \sum_{i=1}^{19} Z_{thrsh}^i$ pixel-wise. Let call the resultant averaged image $Z_{ensemble}$.

3) *Compute confidence:* We computed the confidence score f of a predicted masks using corresponding ensemble mask $Z_{ensemble}$ by adding all the pixels together then dividing by the total number of non-zero pixels t . $f = \frac{1}{t} \sum_{i=1}^m \sum_{j=1}^n Z_{ensemble}^{i,j}$, where m and n represents the dimensions of $Z_{ensemble}$.

4) *Human verification:* In the human verification step, the predicted active set mask using model M_{19} which have a confidence score $f \geq 0.9$ are given to the human for verification. Human verification works the same as previously done with ASA masks, where verification of agreement between predicted masks and the manual annotation is performed. If there is an agreement between the predicted mask and the manual annotation, then EDF/mask is accepted and moved to train set. Meanwhile, it gets removed from the active set. If an EDF/mask do not match manual annotation, then it gets rejected, and the EDF remains in the active set.

C. Testing and post-processing

After completing 5 iterations of the active deep learning in as shown in Fig. 2b, testing of model M_{19} on the test set is done as shown in Fig. 2c. It is worth noting that each model M_{19}^j (i.e., deep learning model saved during training at 100^{th} epoch) is a result of training Unet on a different number of training instances where j is the iteration number and can be $j \in \{1, 2, 3, 4, 5\}$. The results of testing on M_{19}^j on an image x_{EDF} is a probability map x_{prop} of the same size as x_{EDF} . Each pixel p of $x_{prop} \in (0, 1)$ which represent the likelihood of p being either part of a cell (foreground) or not part of a cell (background) based on a given threshold. We thresholded the probability maps x_{prop} at 0.9 because we need the most certain pixels p that belongs to a cell by applying more restrictions on the decision of foreground and background (i.e., cell or not a cell). For instance, $p \geq 0.9$ belong to a cell, then $(p = 255)$, otherwise p is background $(p = 0)$. After thresholding, a post-processing step was applied for three purposes: 1) to remove small noise (blobs) in the predicted mask by removing cells of area size ≤ 250 pixels, 2) apply unbiased stereology counting rules, by removing cells that touch exclusion lines, and 3) separate touching cells by applying the watershed algorithm. Then, the automatic counting step is done to count the total number of cells based on the unbiased stereology counting rules [27].

VII. EXPERIMENTS AND RESULTS

Our data set has 966 NeuN single stain stacks from 9 different mice. The EDF algorithm was used to create an in-focus image for each stack. The number of images in the initial train set (no augmentation) is 147 images, the number of images in the initial active set is 728 images, and the number of images in the test set is 91 images. The data augmentation used in this experiment was a combination of rotation and elastic transformation [32], where the total number of images generated by applying elastic then rotation augmentation of a single image is 72 images (including original image). When testing deep learning models on the unseen test set, we used the model M_{19} which is the last model saved from snapshots models (epoch 100), whereas the ensemble of all snapshot models $M_1, M_2, \dots, M_{19}$ was used only to compute the confidence score f of each predicted masks of the Active set. We have used the error rate to report results on the test set as shown in Equation 1, where y_{true} is the number of counted cells on ground truth (manual annotation), and y_{pred} is the number of counted cells on a predicted deep learning mask. From all 5 iteration models, the best result of active deep learning approach is 0.27% error rate and 0.905 dice coefficient as shown in Table II. Dice coefficient was calculated using Equation 2, where A is the ground truth mask, and B is the predicted mask.

$$Error\ rate = \frac{|y_{true} - y_{pred}|}{y_{true}} * 100 \quad (1)$$

$$Dice = \frac{2 * |A \cap B|}{|A| + |B|} \quad (2)$$

In Table II, a comparison between Iterative deep learning and Active deep learning in terms of a number of images verified, the number of images accepted, and time spent by a human in verification from all five different iterations. The total number of verified images by the human over all the five iterations when using Iterative deep learning approach is 1761 images, which took approximately 101.84 minutes, whereas the total number of verified images by the human across all five iterations when using Active deep learning is 1321 images, which took about 76.39 minutes, where the time taken to verify each mask is approximately 3.47 seconds. Active deep learning proposed in this paper shows a reduction on verification time by approximately 25%, where the human was allowed to verify only the most likely confident images based on confidence score calculated using snapshot ensemble as described in Section VI.

VIII. DISCUSSION

Prior studies that have noted the usefulness of using existing algorithms to generate masks for ground truth [23][33]. However, a human-in-the-loop verification step is required to reject mislabeled images. The drawback of such a verification step is that a human needs to verify the entire set of predicted masks, and thus requires time and effort given the enormous unlabeled data sets.

Labeling data manually especially creating masks is a time-consuming process since it requires pixel-level labeling. Both Iterative deep learning and active deep learning approaches require minimal human intervention, where human is required only to either accept or reject based on the agreement of predicted mask and its corresponding manual annotation. However, the active deep learning study herein found that utilizing the snapshot ensemble approach to generate a confidence score of a predicted mask can reduce the verification time further compared to iterative deep learning since the user is getting only high confident masks to verify. Therefore, the total time consumed in verification using active deep learning approach is 76.39 minutes to verify 1321 masks, whereas iterative deep learning took 101.84 minutes to verify 1761 masks, where verification of each image takes approximately 3.47 seconds. Whereas the results of both iterative deep learning and active deep learning on the test mouse was mostly similar with slightly improvement in active deep learning. A comparison between manual, ASA, iterative deep learning, and active deep learning cell counting per section of the test mouse is shown in Fig. 3. Additionally, a comparison between ASA and active deep learning cell segmentation and counting cells on images from test set are shown in Fig. 4, where best model counted cells contours are overlaid on top of the manual annotation.

This study was subject to some limitations. For instance, the ASA approach works on EDF images to create the initial ground truth masks; however, EDF images may produced obscured cells and create an overlapping of cells in dense areas, and therefore, fewer images were accepted. Additionally, EDF images was also an obstacle in some dense area for deep learning; therefore postprocessing was not able to

separate some overlapping cells. Another limitation of this study is the subjectivity of human verification, which may affect the training of images. Nevertheless, the findings of this study provide an insight into utilizing knowledge driven from snapshot ensembles to reduce the human verification effort and time and thus improving the performance of segmentation and counting of cells using unbiased stereology counting rules.

TABLE II

THE TOTAL NUMBER OF VERIFIED IMAGES, TOTAL NUMBER OF ACCEPTED IMAGES, AND TOTAL VERIFICATION TIME FOR EACH OF ACTIVE DEEP LEARNING AND ITERATIVE DEEP LEARNING

Compare/Deep learning approaches	Iterative Deep Learning	Active Deep Learning
Total verified images by human	1761	1321
Total accepted images by human	544	401
Total verification time (minutes)	101.84	76.39

TABLE III

BEST PERFORMANCE OF DEEP LEARNING (BASELINE), ACTIVE DEEP LEARNING VS. ITERATIVE DEEP LEARNING

Deep learning approach	Error rate (%)	Dice Coef
Deep learning (baseline) [4]	3.57	0.906
Iterative deep learning [23]	0.41	0.912
Active deep learning (Proposed)	0.27	0.905

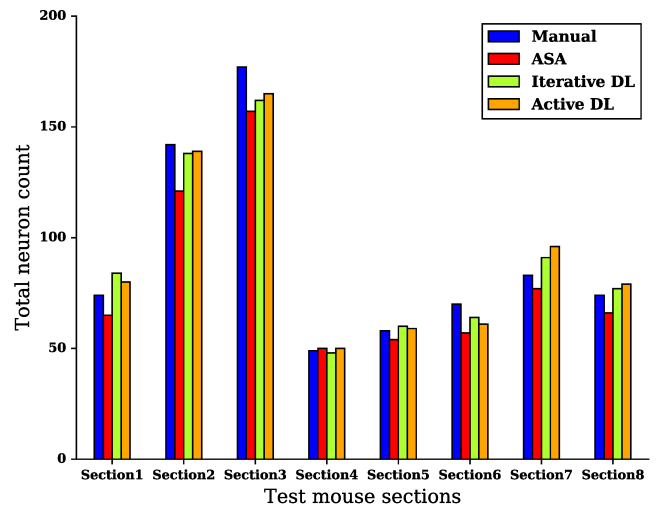


Fig. 3. Test mouse cells count using manual, ASA, and Unet (active deep learning)

IX. CONCLUSIONS

This paper presents an active deep learning algorithm for unbiased stereology cell count that uses a previously existing unsupervised algorithm called ASA to generate initial masks for training deep convolution neural network for

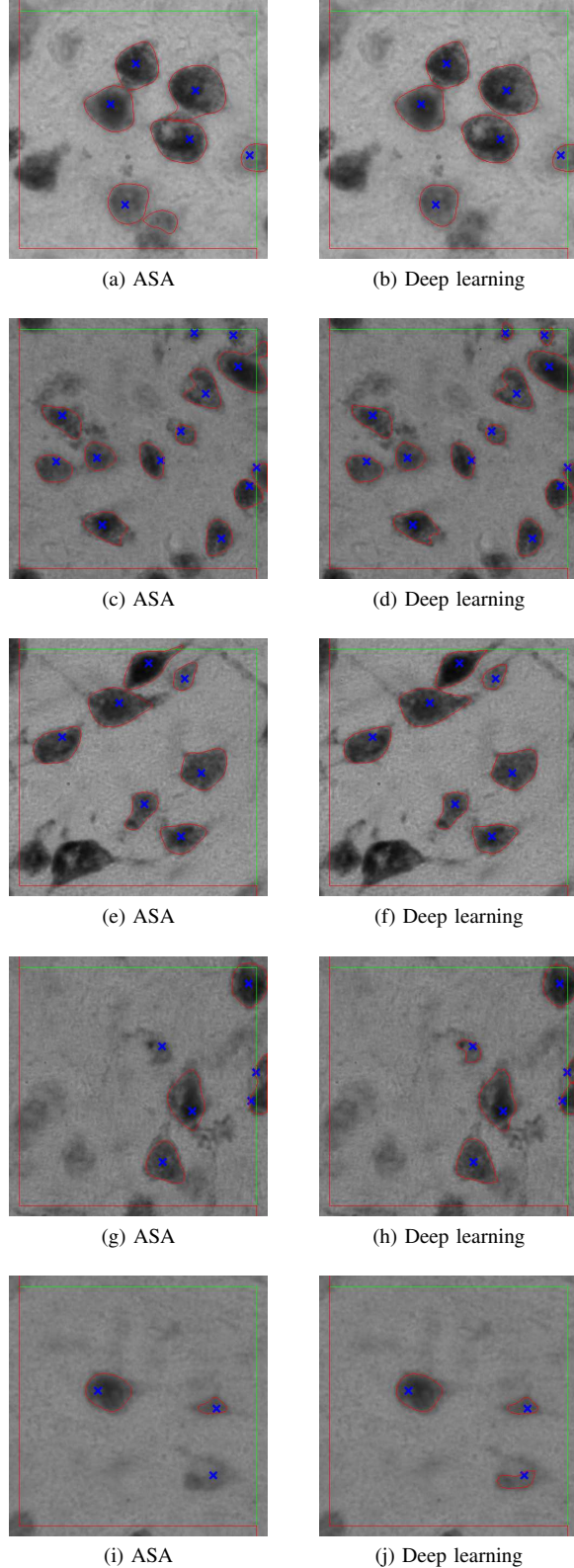


Fig. 4. Examples from the test set, where a,c,e,g and i are the ASA mask contours overlaid on manual annotation images (counted neurons have blue marks), b, d, f, h and j are the Active deep learning predicted masks (iteration 5) contours overlaid on manual annotation image

image segmentation. The proposed method uses the snapshot ensemble to compute a confidence score for each unlabeled EDF image in the active set, where the user gets only the images of higher confidence score to verify, and thus, human-in-the-loop effort in verification was reduced compared to the iterative deep learning approach. The proposed method herein showed a lower error rate of less than 1% compared to the ASA cell counting (an error rate of 11%) on an unseen test mouse images. Additionally, the time consumed by the human-in-the-loop verification was about 25% lower compared to the iterative deep learning approach.

ACKNOWLEDGMENT

The authors would like to thank Dr. Marcia Gordon of Michigan State University (Grand Rapids, MI) for the generous donation of the stained tissue sections for these studies. This research was partially supported by the National Science Foundation under award number (#1513126) and (#1746511).

REFERENCES

- [1] P. R. Mouton, *Unbiased stereology: a concise guide*. JHU Press, 2011.
- [2] P. R. Mouton, H. A. Phoulady, D. Goldgof, L. O. Hall, M. Gordon, and D. Morgan, "Unbiased estimation of cell number using the automatic optical fractionator," *Journal of chemical neuroanatomy*, vol. 80, pp. A1–A8, 2017.
- [3] M. Burke, S. Zangenehpour, P. R. Mouton, and M. Ptito, "Knowing what counts: unbiased stereology in the non-human primate brain," *Journal of visualized experiments: JoVE*, p. 27, 2009.
- [4] S. S. Alahmari, D. Goldgof, L. Hall, H. A. Phoulady, R. H. Patel, and P. R. Mouton, "Automated cell counts on tissue sections by deep learning and unbiased stereology," *Journal of chemical neuroanatomy*, vol. 96, pp. 94–101, 2019.
- [5] A. G. Valdecasas, D. Marshall, J. M. Becerra, and J. Terrero, "On the extended depth of focus algorithms for bright field microscopy," *Micron*, vol. 32, no. 6, pp. 559–569, 2001.
- [6] A. P. Bradley and P. C. Bamford, "A one-pass extended depth of field algorithm based on the over-complete discrete wavelet transform," in *Image and Vision Computing'04 New Zealand (IVCNZ'04)*. not found, 2004, pp. 279–284.
- [7] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [8] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.
- [9] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [10] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.
- [12] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [15] B. Settles, "Active learning literature survey," University of Wisconsin-Madison Department of Computer Sciences, Tech. Rep., 2009.

- [16] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proceedings of the ninth ACM international conference on Multimedia*. ACM, 2001, pp. 107–118.
- [17] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *Journal of machine learning research*, vol. 2, no. Nov, pp. 45–66, 2001.
- [18] Y. Liu, "Active learning with support vector machine applied to gene expression data for cancer classification," *Journal of chemical information and computer sciences*, vol. 44, no. 6, pp. 1936–1941, 2004.
- [19] A. Top, G. Hamarneh, and R. Abugharbieh, "Active learning for interactive 3d image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2011, pp. 603–610.
- [20] C. Sommer, C. Straehle, U. Koethe, and F. A. Hamprecht, "Ilastik: Interactive learning and segmentation toolkit," in *2011 IEEE international symposium on biomedical imaging: From nano to macro*. IEEE, 2011, pp. 230–233.
- [21] A. Vezhnevets, J. M. Buhmann, and V. Ferrari, "Active learning for semantic segmentation with expected change," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 3162–3169.
- [22] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, "Snapshot ensembles: Train 1, get m for free," *arXiv preprint arXiv:1704.00109*, 2017.
- [23] S. Alahmari, D. Goldgof, L. Hall, P. Dave, H. A. Phoulady, and P. Mouton, "Iterative deep learning based unbiased stereology with human-in-the-loop," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 665–670.
- [24] C. B. Saper, "Any way you cut it: a new journal policy for the use of unbiased counting methods," *Journal of Comparative Neurology*, vol. 364, no. 1, pp. 5–5, 1996.
- [25] K. Santacruz, J. Lewis, T. Spires, J. Paulson, L. Kotilinek, M. Ingels-son, A. Guimaraes, M. Deture, M. Ramsden, E. McGowan *et al.*, "Tau suppression in a neurodegenerative mouse model improves memory function," *Science*, vol. 309, no. 5733, pp. 476–481, 2005.
- [26] T. L. Spires, J. D. Orne, K. SantaCruz, R. Pitstick, G. A. Carlson, K. H. Ashe, and B. T. Hyman, "Region-specific dissociation of neuronal loss and neurofibrillary pathology in a mouse model of tauopathy," *The American journal of pathology*, vol. 168, no. 5, pp. 1598–1607, 2006.
- [27] M. West, L. Slomianka, and H. J. G. Gundersen, "Unbiased stereological estimation of the total number of neurons in the subdivisions of the rat hippocampus using the optical fractionator," *The Anatomical Record*, vol. 231, no. 4, pp. 482–497, 1991.
- [28] A. Savitzky and M. J. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [29] F. Chollet *et al.*, "Keras," <https://github.com/fchollet/keras>, 2015.
- [30] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.
- [31] D. Kinga and J. B. Adam, "A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.
- [32] P. Y. Simard, D. Steinkraus, J. C. Platt *et al.*, "Best practices for convolutional neural networks applied to visual document analysis," in *ICDAR*, vol. 3, 2003, pp. 958–962.
- [33] H. A. Phoulady, D. Goldgof, L. O. Hall, and P. R. Mouton, "Automatic ground truth for deep learning stereology of immunostained neurons and microglia in mouse neocortex," *Journal of chemical neuroanatomy*, 2019.